

CERTIFIED COPY OF
PRIORITY DOCUMENT

BEST AVAILABLE COPY

本 国 特 許 庁
JAPAN PATENT OFFICE

Jc872 U.S. PTO
10/039546



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年12月 1日

出 願 番 号

Application Number:

特願2000-366729

出 願 人

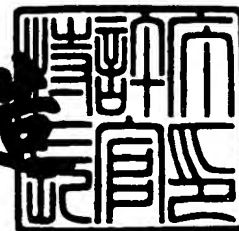
Applicant(s):

株式会社ソニー・コンピュータエンタテインメント

2001年 7月 2日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



JAPAN PATENT OFFICE

#J
1c872 U.S. PRO
10/039546
10/22/01

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: December 1, 2000

Application Number: Patent Application
No. 2000-366729

Applicant(s): Sony Computer Entertainment Inc.

July 2, 2001

Commissioner,
Japan Patent Office

Kouzou OIKAWA

Number of Certificate: 2001-3061153

【書類名】 特許願

【整理番号】 SCEI99209

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明者】

 【住所又は居所】 東京都港区赤坂 7 丁目 1 番 1 号 株式会社ソニー・コンピュータエンタテインメント内

 【氏名】 馬越 英尚

【発明者】

 【住所又は居所】 東京都港区赤坂 7 丁目 1 番 1 号 株式会社ソニー・コンピュータエンタテインメント内

 【氏名】 佐々木 伸夫

【特許出願人】

 【識別番号】 395015319

 【氏名又は名称】 株式会社ソニー・コンピュータエンタテインメント

【代理人】

 【識別番号】 100107238

 【弁理士】

 【氏名又は名称】 米山 尚志

【手数料の表示】

 【予納台帳番号】 111236

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 データベース構築方法及び装置、処理プログラムを情報処理装置に実行させる媒体

【特許請求の範囲】

【請求項 1】 ソース情報から文字情報を抽出し、
上記文字情報から単語情報を抽出し、
上記単語情報によりデータベースを構築する
ことを特徴とするデータベース構築方法。

【請求項 2】 所定の通信ネットワーク上で検索及び／又は閲覧されたページを上記ソース情報として取り込むことを特徴とする請求項 1 記載のデータベース構築方法。

【請求項 3】 電子メール情報を上記ソース情報として取り込むことを特徴とする請求項 1 記載のデータベース構築方法。

【請求項 4】 画像信号を上記ソース情報として取り込み、
上記画像信号から文字認識により上記文字情報を抽出することを特徴とする請求項 1 記載のデータベース構築方法。

【請求項 5】 記憶装置或いは記憶媒体内に保存されているデータファイルを上記ソース情報として取り込むことを特徴とする請求項 1 記載のデータベース構築方法。

【請求項 6】 上記ソース情報から抽出した各文字情報に順次個別にフラグを設定することを特徴とする請求項 1 から請求項 5 のうち、いずれか一項記載のデータベース構築方法。

【請求項 7】 既に設定された上記フラグに応じて、ソース情報からの文字情報の抽出の可否を判定することを特徴とする請求項 6 記載のデータベース構築方法。

【請求項 8】 上記抽出した単語情報を、既存のデータベース内の単語情報と照合し、

上記抽出した単語情報が上記既存のデータベース内に存在しないとき、当該抽出した単語情報を上記既存のデータベースに追加登録することを特徴とする請求

項 1 から請求項 7 のうち、何れか一項記載のデータベース構築方法。

【請求項 9】 上記抽出した単語情報が上記既存のデータベース内に存在するとき、さらに上記既存のデータベース上で最高優先順位が設定された単語情報と同一であるか判定し、

上記抽出した単語情報が上記最高優先順位の単語情報と同一でないとき、上記既存のデータベースを上記抽出した単語情報により更新することを特徴とする請求項 8 記載のデータベース構築方法。

【請求項 10】 既存のデータベースの変更の可否を判定し、

上記既存のデータベースの変更が許可されているときにのみ当該既存のデータベースの変更を行うことを特徴とする請求項 8 又は請求項 9 記載のデータベース構築方法。

【請求項 11】 上記既存のデータベースの変更の際に、変更確認の設定の有無を判定し、

上記変更確認の設定がなされているときには所定の確認メッセージを発生することを特徴とする請求項 10 記載のデータベース構築方法。

【請求項 12】 上記抽出した単語情報の読みと当該単語情報とを対応付けて上記データベースを構築することを特徴とする請求項 1 から請求項 11 のうち、いずれか一項記載のデータベース構築方法。

【請求項 13】 ソース情報から文字情報を抽出する文字情報抽出部と、
上記文字情報から単語情報を抽出する単語情報抽出部と、
上記単語情報によりデータベースを構築する構築部とを有することを特徴とするデータベース構築装置。

【請求項 14】 所定の通信ネットワークと接続する接続部と、
上記所定の通信ネットワーク上で検索及び／又は閲覧されたページを上記ソース情報として取り込むソース情報取り込み部とを有することを特徴とする請求項 13 記載のデータベース構築装置。

【請求項 15】 電子メール情報を上記ソース情報として取り込むソース情報取り込み部を有することを特徴とする請求項 13 記載のデータベース構築装置。

【請求項 1 6】 画像信号を上記ソース情報として取り込むソース情報取り込み部を有し、

上記文字情報抽出部は、上記画像信号から文字認識により上記文字情報を抽出することを特徴とする請求項 1 3 記載のデータベース構築装置。

【請求項 1 7】 記憶装置或いは記憶媒体内に保存されているデータファイルを上記ソース情報として取り込むソース情報取り込み部を有することを特徴とする請求項 1 3 記載のデータベース構築装置。

【請求項 1 8】 上記ソース情報から抽出した各文字情報に順次個別にフラグを設定するフラグ設定部を有することを特徴とする請求項 1 3 から請求項 1 7 のうち、いずれか一項記載のデータベース構築装置。

【請求項 1 9】 既に設定された上記フラグに応じて、ソース情報からの文字情報の抽出の要否を判定する判定部を有することを特徴とする請求項 1 8 記載のデータベース構築装置。

【請求項 2 0】 上記構築部は、上記抽出した単語情報を、既存のデータベース内の単語情報と照合し、上記抽出した単語情報が上記既存のデータベース内に存在しないとき、当該抽出した単語情報を上記既存のデータベースに追加登録することを特徴とする請求項 1 3 から請求項 1 9 のうち、何れか一項記載のデータベース構築装置。

【請求項 2 1】 上記構築部は、上記抽出した単語情報が上記既存のデータベース内に存在するとき、さらに上記既存のデータベース上で最高優先順位が設定された単語情報と同一であるか判定し、上記抽出した単語情報が上記最高優先順位の単語情報と同一でないとき、上記既存のデータベースを上記抽出した単語情報により更新することを特徴とする請求項 2 0 記載のデータベース構築装置。

【請求項 2 2】 上記構築部は、既存のデータベースの変更の可否を判定し、上記既存のデータベースの変更が許可されているときにのみ当該既存のデータベースの変更を行うことを特徴とする請求項 2 0 又は請求項 2 1 記載のデータベース構築装置。

【請求項 2 3】 上記既存のデータベースの変更の際に、変更確認の設定の有無を判定し、上記変更確認の設定がなされているときには所定の確認メッセー

ジを発生する変更確認部を有することを特徴とする請求項22記載のデータベース構築装置。

【請求項24】 上記構築部は、上記抽出した単語情報の読みと当該単語情報とを対応付けて上記データベースを構築することを特徴とする請求項13から請求項23のうち、いずれか一項記載のデータベース構築装置。

【請求項25】 ソース情報から文字情報を抽出するステップと、
上記文字情報から単語情報を抽出するステップと、
上記単語情報によりデータベースを構築するステップとを含むことを特徴する処理プログラムを情報処理装置に実行させる媒体。

【請求項26】 所定の通信ネットワーク上で検索及び／又は閲覧されたページを上記ソース情報として取り込むステップを含むことを特徴とする請求項25記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項27】 電子メール情報を上記ソース情報として取り込むステップを含むことを特徴とする請求項25記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項28】 画像信号を上記ソース情報として取り込むステップと、
上記画像信号から文字認識により上記文字情報を抽出するステップとを含むことを特徴とする請求項25記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項29】 記憶装置或いは記憶媒体内に保存されているデータファイルを上記ソース情報として取り込むステップを含むことを特徴とする請求項25記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項30】 上記ソース情報から抽出した各文字情報に順次個別にフラグを設定するステップを含むことを特徴とする請求項25から請求項29のうち、いずれか一項記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項31】 既に設定された上記フラグに応じて、ソース情報からの文字情報の抽出の可否を判定するステップを含むことを特徴とする請求項30記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項32】 上記抽出した単語情報を、既存のデータベース内の単語情

報と照合するステップと、

上記抽出した単語情報が上記既存のデータベース内に存在しないとき、当該抽出した単語情報を上記既存のデータベースに追加登録するステップとを含むことを特徴とする請求項 2 5 から請求項 3 1 のうち、何れか一項記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項 3 3】 上記抽出した単語情報が上記既存のデータベース内に存在するとき、さらに上記既存のデータベース上で最高優先順位が設定された単語情報と同一であるか判定するステップと、

上記抽出した単語情報が上記最高優先順位の単語情報と同一でないとき、上記既存のデータベースを上記抽出した単語情報により更新するステップとを含むことを特徴とする請求項 3 2 記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項 3 4】 既存のデータベースの変更の可否を判定するステップと、

上記既存のデータベースの変更が許可されているときにのみ当該既存のデータベースの変更を行うステップとを含むことを特徴とする請求項 3 2 又は請求項 3 3 記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項 3 5】 上記既存のデータベースの変更の際に、変更確認の設定の有無を判定するステップと、

上記変更確認の設定がなされているときには所定の確認メッセージを発生するステップとを含むことを特徴とする請求項 3 4 記載の処理プログラムを情報処理装置に実行させる媒体。

【請求項 3 6】 上記抽出した単語情報の読みと当該単語情報とを対応付けて上記データベースを構築するステップを含むことを特徴とする請求項 2 5 から請求項 3 5 のうち、いずれか一項記載の処理プログラムを情報処理装置に実行させる媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、例えば仮名漢字、ローマ字漢字変換システムに使用される辞書等の

データベースを構築する場合に好適なデータベース構築方法及び装置、データベース構築のための処理プログラムを情報処理装置に実行させる媒体に関する。

【0002】

【従来の技術】

従来より、例えばパーソナルコンピュータや文書作成装置（いわゆるワードプロセッサ）を用いた仮名漢字、ローマ字漢字変換システム（以下、適宜、漢字変換システムと呼ぶ）では、例えばキーボード等からの入力信号を仮名或いはローマ字コードに変換し、それらのコードに対応する仮名或いはローマ字の文字画像を表示装置の画面上に表示したり、さらに一又は複数の連続した仮名或いはローマ字コード列を漢字コードに変換し、それら漢字コードに対応する漢字文字画像を表示するようなことが行われている。なお、仮名漢字変換とは仮名コードを漢字コードに変換することであり、ローマ字漢字変換とはローマ字コードを漢字コードに変換することである。

【0003】

また、上記漢字変換システムは、一般に、仮名と単語との対応表を蓄積している辞書データベースを備えており、自動的若しくはユーザの指示に応じて、一又は複数の連続した仮名から単語への変換を行うようにもなされている。なお、上記単語とは、複数の漢字の組み合わせや、漢字と仮名の組み合わせ、複数のカタカナ等からなるものである。

【0004】

【発明が解決しようとする課題】

ところで、上記漢字変換システムに予め用意されているオリジナルの辞書データベースは、例えば仮名漢字、ローマ字漢字変換ソフトウェアベンダ等が作成及び構築したものである。また、当該辞書データベースは、例えばユーザの使用効率等を考慮して、新たな単語の追加及び更新や、単語の使用頻度に応じた変換優先順位の自動設定などが可能となされていることが多い。

【0005】

しかしながら、上記辞書データベースに意図的に新たな単語を追加するためには、例えばユーザ自らが辞書データベース内から所望の漢字や仮名を探し出し、

それらを組み合わせて単語を生成し、さらにその単語を辞書データベースへ追加登録する指示、或いは辞書データベースの更新指示を与えるような作業が必要になる。

【0006】

また、上記変換優先順位を意図的に変更するためには、ユーザ自らが上記仮名から単語への変換のための入力操作を例えば複数回繰り返す等の作業を行わなければならない。

【0007】

以上のように、従来の漢字変換システムにおいて、新たな単語の追加及び更新や変換優先順位の意図的な設定を行うためには、ユーザの手作業による幾つもの煩雑な操作が必要となり、非常に効率が悪い。

【0008】

そこで、本発明は、このような課題に鑑みてなされたものであり、例えば、漢字変換システムの辞書データベースに対する新たな単語の追加及び更新や変換優先順位の設定を、ユーザによる複雑な操作を必要とせず、簡単且つ容易に実現可能とする、データベース構築方法及び装置と、上記データベースの構築を実現する処理プログラムを情報処理装置に実行させる媒体を提供することを目的とする。

【0009】

【課題を解決するための手段】

本発明では、ソース情報から文字情報を抽出し、その文字情報から単語情報を抽出し、単語情報によりデータベースを構築する。上記ソース情報は、所定の通信ネットワーク上で検索及び／又は閲覧されたページや、電子メール情報、画像信号などである。

【0010】

さらに、本発明では、ソース情報から抽出した各文字情報に順次個別にフラグを設定し、既に設定された上記フラグに応じて、ソース情報からの文字情報の抽出の要否を判定する。また、抽出した単語情報を、既存のデータベース内の単語情報と照合し、既存のデータベース内に存在しないときには、当該抽出した単語

情報を上記既存のデータベースに追加登録し、一方、既存のデータベース内に存在するときには、上記抽出した単語情報が上記既存のデータベース上で最高優先順位の単語情報と同一でないときにのみその抽出した単語情報によって既存のデータベースを更新する。

【 0 0 1 1 】

すなわち、本発明によれば、ソース情報は、例えば使用者が所定の通信ネットワーク上で検索及び／又は閲覧しているページや使用者がやりとりしている電子メール情報等であり、そのソース情報から文字情報及び単語情報を抽出し、その抽出した単語情報によりデータベースを構築するようにしている。このため、当該構築されたデータベースは使用者の専門分野或いは興味のある分野に適合したものとなる。また、データベースの構築は自動的或いは半自動的に行われるため、使用者は当該データベースの構築のために、手作業による単語の追加や登録などの作業を行う必要はない。

【 0 0 1 2 】

【発明の実施の形態】

〔本発明の実施の形態の概要〕

本発明実施の形態の辞書データベース構築装置は、例えば仮名漢字、ローマ字漢字変換システムなどで使用される辞書データベースを構築するための装置であり、例えば外部より取り込んだソース情報から文字情報を抽出し、さらにその文字情報から単語を抽出し、それら抽出した単語を用いて辞書データベースの更新、追加登録などを行うことにより、使い易い辞書データベースを簡単且つ容易に構築可能とするものである。

【 0 0 1 3 】

ここで、例えばユーザ自らがインターネット検索、閲覧や電子メールのやりとりを行って得られた文字情報や、テレビジョン信号やビデオ信号などの画像信号、或いは本や雑誌などをスキャニングして取り込まれた画像信号から文字認識等の手法により得られた文字情報、CD-ROMやDVD-ROM等の各種の記録媒体から得られた文字情報などには、ユーザにとって興味のある分野、或いはユーザの専門分野に関する単語が多数含まれていると考えられる。また、ユーザが

仮名漢字，ローマ字漢字変換システムを用いて例えば文書等を作成する際には、それら興味のある分野、或いは専門分野に関する単語を使用する可能性が高い（又は使用頻度が高い）と考えられる。

【0014】

このようなことから、本発明の実施の形態の辞書データベース構築装置では、上記インターネット検索、閲覧、電子メールによる文字情報や、テレビジョン信号等やスキャニングによる画像信号から得られた文字情報、上記CD-ROM等から得られた文字情報などから単語を抽出し、例えば仮名漢字，ローマ字漢字変換システムに予め用意されていたオリジナルの辞書データベースに、上記抽出した単語を追加登録及び更新し、また、当該抽出した単語についての変換優先順位を上げるなどの処理を行うようにしている。

【0015】

すなわち、本発明実施の形態によれば、上記抽出した単語を辞書データベースに追加登録及び更新したり、上記抽出した単語について変換優先順位の設定を行うことにより、従来のようなユーザの手作業による複雑な操作を必要とせず、簡単且つ容易に、使い易い辞書データベースを構築可能としている。また、本発明実施の形態によれば、特に、最近のカタカナ表記による外来語を含んだ辞書データベースの構築の際に大変有効であると考えられる。なお、本実施の形態の辞書データベース構築装置は、既存の辞書データベースに単語を追加登録及び更新する場合に限らず、新規に辞書データベースを構築する場合にも適用可能であることは言うまでもない。

【0016】

〔本実施の形態の概略ブロック構成〕

図1には、本発明の一実施の形態として、辞書データベース構築のための処理プログラム（以下、辞書データベース構築プログラムと呼ぶ）を実行する装置の概略的なブロック構成を示す。なお、この図1には、本実施の形態の辞書データベース構築プログラムを例えばパーソナルコンピュータにより実行することにした場合の主要部の構成のみを示している。

【0017】

図1において、記憶部15は、例えばハードディスク及びそのドライブからなる。上記ハードディスク内には、オペレーティングシステムプログラムの他、本実施の形態の辞書データベース構築処理を実現するための後述する辞書データベース構築プログラムや、例えば仮名漢字、ローマ字漢字変換システムを実現するための仮名漢字、ローマ字漢字変換プログラム、いわゆるインターネットに接続及び閲覧するためのブラウザプログラム、電子メールの作成や送受信を行うためのメールプログラム、ビデオ信号やテレビジョン信号など各種の画像信号を取り込むキャプチャプログラム、例えばフラットヘッドスキャナ等により読みとられた画像データや上記キャプチャされた画像信号などから文字認識を行って文字データへの変換等を行う文字認識プログラムや、CD-ROMやDVD-ROM等の各種の記録媒体からデータを取り込むプログラムなどを含む各種のプログラム16と、その他の各種データ17が記憶されている。また、上記各種データ17には、本実施の形態の辞書データベース18も含まれている。なお、本実施の形態の辞書データベース構築プログラムは、例えば上記仮名漢字、ローマ字漢字変換プログラムなど、他のプログラム内に含まれるものであっても良い。

【0018】

上記辞書データベース18のオリジナル（初期辞書データベース）は、例えば、仮名漢字、ローマ字漢字変換ソフトウェアを作成したベンダ等により提供されたものであり、上述のように新たな単語の追加及び更新や、単語の使用頻度に応じた変換優先順位の自動設定などが可能となされている。なお、上記辞書データベース18は、上記仮名漢字、ローマ字漢字変換ソフトウェアを作成したベンダにより提供されるもの以外に、他の仮名漢字、ローマ字漢字変換ソフトウェアから取り込んだものであっても良く、また、本実施の形態の辞書データベース構築プログラムに付属して提供されたものでも、或いは、辞書データベースのみが単独で提供されたものであっても良い。

【0019】

通信部10は、例えば、アナログ公衆電話回線に接続するためのモデム、ケーブルテレビジョン網に接続するためのケーブルモデム、ISDN（総合デジタル通信網）に接続するためのターミナルアダプタ（Terminal Adaptor：TA）な

どのように、外部とデータ通信を行うための通信デバイスである。通信 I/F 部 11 は、上記通信部 10 と内部バス (BUS) との間でデータのやりとりを可能とするためのプロトコル変換等を行うインターフェイスデバイスである。本実施の形態では、上記通信部 10 及び通信 I/F 部 11 により、例えばいわゆるインターネットへの接続、検索、閲覧、電子メールの送受信等が可能となされている。

【0020】

外部機器 25 は、例えばフラットヘッドスキャナや、TV チューナ、ビデオテープレコーダ、ビデオディスクプレーヤ、デジタルカメラ、ビデオカメラ、衛星放送受信用の BS チューナ、通信衛星回線接続用の CS チューナなどの機器である。外部機器 I/F 部 26 は、上記外部機器 25 から供給された信号を内部に取り込むためのインターフェイスデバイスである。

【0021】

入力部 20 は、例えばキーボードやマウス、タッチパッドなどの入力装置であり、ユーザ I/F 部 19 は、上記入力部 20 からの信号を内部に供給するためのインターフェイスデバイスである。

【0022】

ドライブ部 22 は、例えば CD-ROM や DVD-ROM、フロッピーディスク等のディスク媒体や、カード状等の半導体メモリなどから、少なくともデータを読み出し可能なドライブ装置である。ドライブ I/F 部 21 は、上記ドライブ部 22 からの信号を内部に供給するためのインターフェイスデバイスである。

【0023】

表示部 24 は、例えば CRT (陰極線管) や液晶等の表示デバイスであり、表示ドライブ部 23 は上記表示部 24 を表示駆動させるドライブデバイスである。

【0024】

CPU 12 は、上記記憶部 15 に記憶されているオペレーティングシステムプログラムや上記プログラム 16 に基づいて、図 1 のパーソナルコンピュータの全動作を制御する。

【0025】

ROM 1 3 は、例えばフラッシュメモリ等の書き換え可能な不揮発性メモリからなり、図 1 のパーソナルコンピュータの B I O S (Basic Input/Output System) や各種の初期設定値を記憶している。RAM 1 4 は、記憶部 1 5 のハードディスクから読み出されたアプリケーションプログラムなどがロードされ、また、CPU 1 2 のワーク RAM として用いられる。

【 0 0 2 6 】

この図 1 に示す構成において、CPU 1 2 は、上記記憶部 1 5 のハードディスクから読み出されて RAM 1 4 にロードされたアプリケーションプログラムの一つである、以下の本実施の形態の辞書データベース構築プログラムを実行することにより、辞書データベースの構築、若しくは、辞書データベースへ新たな単語の追加登録及び更新、変換優先順位の設定などの処理を行う。

【 0 0 2 7 】

〔本実施の形態の辞書データベース構築プログラムの概略的な構成〕

〔辞書データベース構築プログラムの基本構成〕

次に、本実施の形態の CPU 1 2 が辞書データベース構築処理を実行するためのアプリケーションプログラム（辞書データベース構築プログラム）の構成を説明する。

【 0 0 2 8 】

本実施の形態の辞書データベース構築プログラムは、例えば DVD-ROM や CD-ROM のような光ディスクや半導体メモリ等の記録媒体に記録されたり、通信回線を介してダウンロード可能なものであり、例えば図 2 に示すようなデータ構成を有している。なお、この図 2 に示すデータ構成は、本実施の形態の辞書データベース構築プログラムの基本構成を概念的に表すものであり、実際のプログラム構成を表しているものではない。

【 0 0 2 9 】

この図 2 に示すように、本実施の形態の辞書データベース構築プログラム 3 0 は、図 1 の CPU 1 2 が実行するプログラム部 3 1 として、文字情報処理プログラム 3 2 と、フラグ管理プログラム 3 3 と、辞書管理プログラム 3 4 と、ユーザ入力情報管理プログラム 3 5 とを少なくとも備えている。

【 0 0 3 0 】

上記文字情報処理プログラム 3 2 は、上記インターネット検索、閲覧ページや電子メールから文字情報を抽出したり、テレビジョン信号等やスキャニングによる画像信号に含まれる文字を文字認識等の手法により抽出したり、上記 C D - R O M 等から得られたデータに含まれる文字情報を抽出し、さらに、それらの文字情報から単語を抽出するためのプログラムである。

【 0 0 3 1 】

上記フラグ管理プログラム 3 3 は、上記抽出された各文字情報に順次個別のフラグを付加することで、例えば重複した文字情報の抽出が行われないようにしたり、或いは、逆に同一の文字情報の重複数を計測して出現頻度を求めるなどのフラグ管理を行うためのプログラムである。なお、本実施の形態では、当該フラグを用いることにより、重複した文字情報の抽出が行われないようにしている。

【 0 0 3 2 】

上記辞書管理プログラム 3 4 は、辞書データベースへの追加登録、更新、単語の使用頻度に応じた変換優先順位の設定などの辞書管理を行うためのプログラムである。

【 0 0 3 3 】

上記ユーザ入力情報管理プログラム 3 5 は、例えば辞書データベースへの追加登録や更新を行うか否かの指示や、変換優先順位の自動設定或いは手動設定の指示がユーザによりなされたとき、その入力情報を管理したり、単語の読みがユーザにより入力されたときに、その読みの情報と単語とを対応つける等のユーザ入力情報を管理するためのプログラムである。

【 0 0 3 4 】

〔仮名漢字，ローマ字漢字変換プログラムへの適用例〕

図 2 の例は、辞書データベース構築プログラム単体の基本構成を示したが、本実施の形態の辞書データベース構築プログラムが例えば仮名漢字，ローマ字漢字変換プログラムに含まれている場合は、例えば図 3 に示すようなデータ構成となる。なお、この図 3 に示すデータ構成も図 2 の例と同様に、アプリケーションプログラムを概念的に表すものであり、実際のプログラム構成を表しているもので

はない。

【 0 0 3 5 】

この図 3 に示す仮名漢字，ローマ字漢字変換アプリケーションプログラム 4 0 は、大別して、漢字変換プログラム部 4 1 と、図 2 に示したプログラム部 3 1 と同様の本実施の形態の辞書データベース構築プログラム 4 2 と、仮名漢字，ローマ字漢字変換に使用されるデータ部 4 3 とを有している。なお、辞書データベース構築プログラム 4 2 については図 2 のプログラム部 3 1 と同様であるため、ここではその説明を省略する。また、上記漢字変換プログラム 4 1 は、いわゆるワープロソフトと呼ばれるような文書作成ソフトウェアと略々同じものである。

【 0 0 3 6 】

上記データ部 4 3 は、仮名漢字，ローマ字漢字変換に使用される仮名コード、ローマ字コード、漢字コード、その他の記号コードからなるコードデータ 4 4 と、仮名と単語の対応表を表す辞書データベース用の辞書データ 4 5 とからなる。ここで、例えば上記 C D - R O M や D V D - R O M 等により提供される辞書データ 4 5、すなわち本実施の形態のパーソナルコンピュータに当該仮名漢字，ローマ字漢字変換アプリケーションプログラム 4 0 がインストールされた直後の辞書データ 4 5 は、オリジナルの初期辞書データベースとして予め用意されたものである。

【 0 0 3 7 】

[本実施の形態の辞書データベース構築プログラム実行時の流れ]

以下、上述の図 2 及び図 3 に示した本実施の形態の辞書データベース構築プログラム 3 0 又は 4 2 における処理の流れを、図 4 に示すフローチャートを用いて説明する。なお、以下に説明する流れは、本実施の形態の辞書データベース構築プログラムを構成する各プログラム部分が、図 1 に示したパーソナルコンピュータの C P U 1 2 上で動作することにより実現されるものである。

【 0 0 3 8 】

この図 4 に示す本実施の形態の辞書データベース構築プログラムにおいて、先ず、文書情報処理プログラム 3 2 は、ステップ S 1 の処理として、インターネット検索、閲覧ページや電子メール、テレビジョン信号やスキャニングによる画像

信号、ハードディスクドライブやCD-ROM、DVD-ROMドライブ、半導体メモリ装置などの各種の記憶装置（或いはそれらに内蔵或いは装填された記憶媒体）に保存されているデータファイル等をソース情報として取り込み、次いで、ステップS2として、それら取り込んだ情報から文字情報を抽出する。なお、例えばインターネット検索、閲覧ページの場合は、HTML（Hyper Text Markup Language）ファイルのテキスト領域から文字情報を抽出し、例えば画像信号の場合は、例えば文字認識処理等によって文字情報を抽出する。また、上記記憶装置或いは記憶媒体に保存されているデータファイルは、ユーザが明示して指定しても、或いはしなくても（すなわち任意に検索）良い。

【0039】

次に、ステップS3として、フラグ管理プログラム33は、上記抽出した各文字情報にフラグを付加する。次に、フラグ管理プログラム33は、ステップS4として、上記各文字情報について上記フラグに「1」が立てられているか否か判定し、フラグに「1」が立てられている文字情報についてはステップS14として破棄する。なお、本実施の形態において、このように重複した文字情報を破棄するようにしたのは、後のステップS5以降の処理を軽減するためである。

【0040】

一方、フラグに「1」が立てられていない（フラグが「0」）文字情報列について、文字情報処理プログラム32は、ステップS5として、その文字情報列を単語毎に分解し、各単語を抽出する。

【0041】

次に、辞書管理プログラム34は、ステップS6として、上記抽出した単語が、既存の辞書データベース上に存在しているか否か照合する。このステップS6において、上記単語が既存の辞書データベース上に存在していないと判定した場合、辞書管理プログラム34は、ステップS7として、当該単語を辞書データベースに追加登録する単語に設定する。一方、上記既存の辞書データベース上に存在している単語の場合、辞書管理プログラム34は、ステップS8として、既存の辞書データベース内で変換最高優先順位となっていない単語のみ残し、その単語で既存の辞書データベースを更新する。なお、このように変換最高優先順位で

ない単語のみを残すようにするのは、既にユーザの使用によって変換最高優先順位となっている単語が、上記抽出された単語により変更されてしまうことを防止するためである。

【0042】

その後、ステップS9として、辞書管理プログラム34は、ユーザ入力情報管理プログラム35が管理しているユーザ入力情報を元に、ユーザが辞書データベースの変更を包括的に許可しているか否かの判定を行う。ここで、許可していると判定された場合はステップS10の処理に進み、許可していないと判定された場合はステップS11の処理に進む。

【0043】

ステップS10の処理に進むと、辞書管理プログラム34は、上記単語を用いて辞書データベースを更新する。

【0044】

これに対し、ステップS11の処理に進んだ場合、ユーザ入力情報管理プログラム35は、例えば上記表示部24上に更新の許可を求める確認メッセージを表示させるような処理を行い、その後、ステップS12として、ユーザから更新許可がなされたか否かを判定する。ステップS12において、更新許可がユーザによりなされた場合はステップS13の処理に進み、許可がなされない場合はステップS14にて当該単語を破棄する。

【0045】

上記ステップS13に進むと、ユーザ入力情報管理プログラム35は、例えば当該単語の「読み」などの入力がユーザによりなされたか否かの判定を行う。このステップS13において、ユーザ入力がない場合はステップS10に進み、ユーザ入力がない場合はステップS15に進む。

【0046】

ステップS13からステップS10に進んだ場合、辞書管理プログラム34はその単語を用いて辞書データベースを更新する。

【0047】

また、ステップS15に進んだ場合、辞書管理プログラム34は、上記ユーザ

入力情報管理プログラム35から渡された「読み」等の入力データと上記単語とを関連付けて、辞書データベースの更新を行う。

【0048】

【実施の形態のまとめ】

以上のように、本実施の形態によれば、ユーザにとって興味のある分野、或いはユーザの専門分野に関する単語を含む辞書データベースを、簡単且つ容易に生成することができる。また例えば本実施の形態を仮名漢字、ローマ字漢字変換システムに適用すれば、非常に使い易いシステムを実現できる。

【0049】

なお、本発明は例えばパーソナルコンピュータに適用した例を挙げているが、例えば文書作成機能を備えた携帯電話機などの携帯型情報端末や、通信或いは情報入力機能を備えた文書作成装置（いわゆるワードプロセッサ）、通信或いは情報入力機能と文書作成機能を備えた家庭用ゲーム機などにも適用可能である。また、上述の説明では、主に仮名漢字、ローマ字漢字変換システムに本発明を適用した例を挙げているが、本発明は、漢字変換システムそのものではなく、辞書データベースの構築のみ行うシステムであっても良いことは言うまでもない。

【0050】

さらに上記実施の形態では、日本語の仮名漢字、ローマ字から漢字へのローマ字漢字変換を例に挙げているが、本発明は他の言語、例えば中国語におけるローマ字漢字変換システムや、ローマ字ハングル文字変換システム、ローマ字アラビア文字変換システムなどにも適用可能である。また、本発明は、英語、フランス語、ドイツ語等の単語辞書データベースの構築やスペルチェック用辞書データベースの構築の際にも適用可能である。特に、本発明を上記スペルチェック用辞書データベースに適用した場合は、当該スペルチェック用辞書データベースにおいて正しいスペル候補に変換する際の変更優先順位の更新などに使用可能である。

【0051】

【発明の効果】

本発明においては、所定の通信ネットワーク上で検索及び／又は閲覧されたページや、電子メール情報、画像信号などのソース情報から文字情報を抽出し、そ

の文字情報から単語情報を抽出し、当該単語情報によりデータベースを構築することにより、例えば、漢字変換システムの辞書データベースに対する新たな単語の追加及び更新や変換優先順位の設定を、ユーザによる複雑な操作を必要とせず、簡単且つ容易に実現可能である。

【図面の簡単な説明】

【図 1】

本発明実施の形態の辞書データベース構築プログラムを実行するパーソナルコンピュータの主要部の構成を示すブロック回路図である。

【図 2】

本実施の形態の辞書データベース構築プログラムの基本構成を概念的に表す図である。

【図 3】

本実施の形態の辞書データベース構築プログラムを含む仮名漢字、ローマ字漢字変換アプリケーションプログラムの構成を概念的に表す図である。

【図 4】

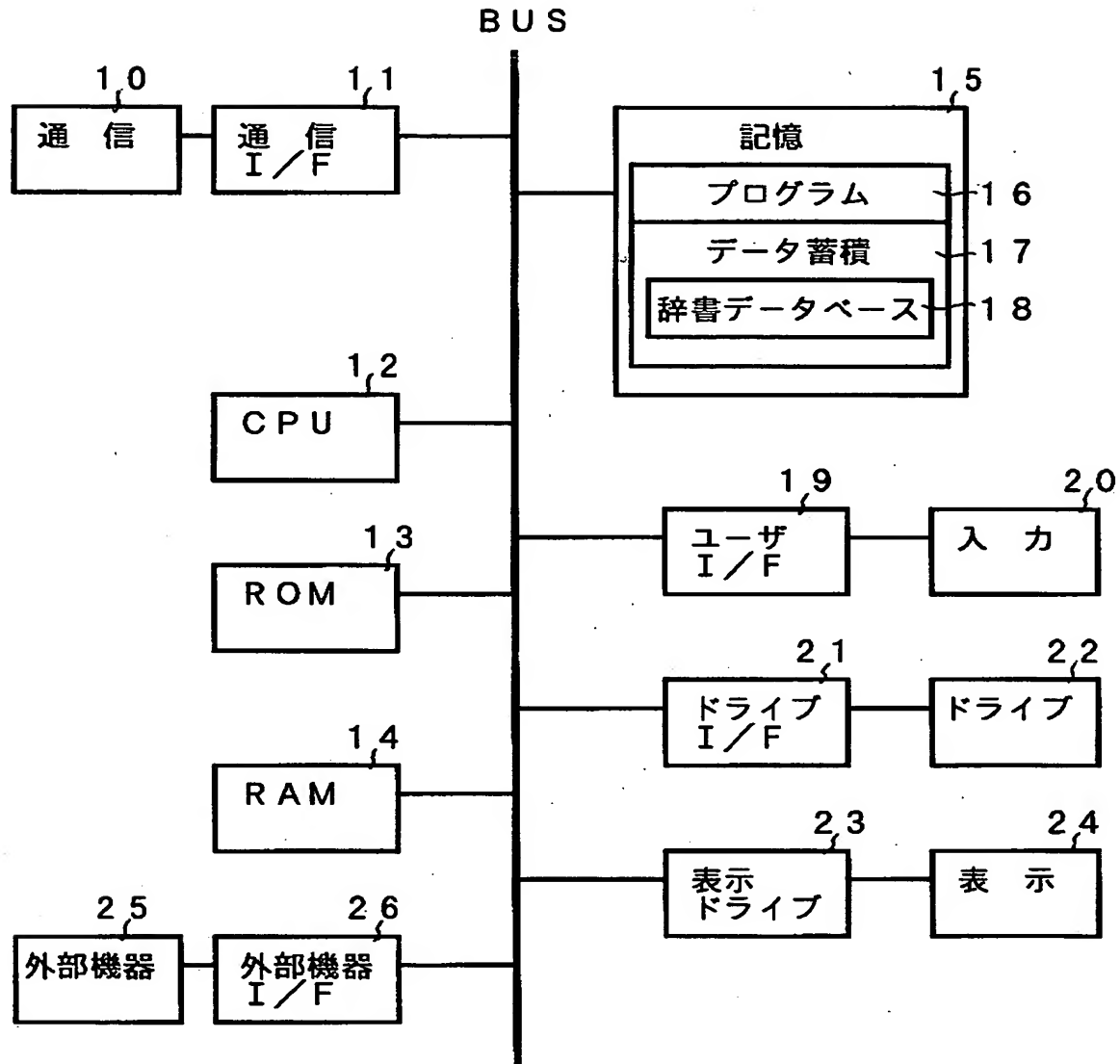
本実施の形態の辞書データベース構築プログラムにおける処理の流れを示すフローチャートである。

【符号の説明】

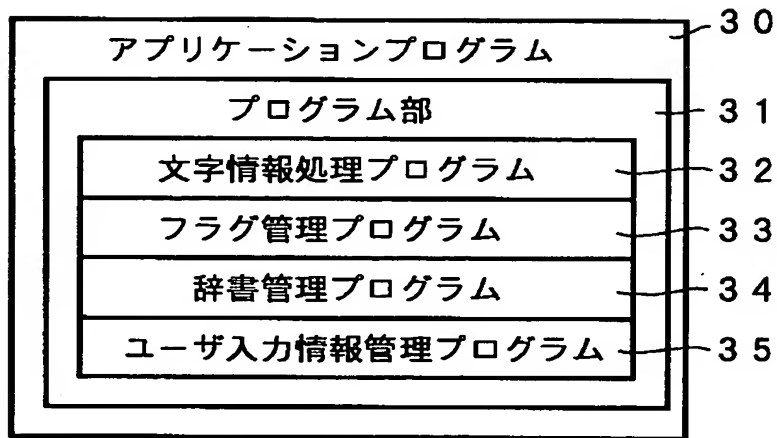
1 0 …通信部、1 1 …通信 I / F 部、1 2 …CPU、1 3 …ROM、1 4 …RAM、1 5 …記憶部、1 6 …プログラム、1 7 …データ蓄積部、1 8 …辞書データベース、1 9 …ユーザ I / F 部、2 0 …入力部、2 1 …ドライブ I / F 部、2 2 …ドライブ部、2 3 …表示ドライブ部、2 4 …表示部、2 5 …外部機器、2 6 …外部機器 I / F 部、3 1, 4 2 …辞書データベース構築プログラム部、3 2 …文字情報処理プログラム、3 3 …フラグ管理プログラム、3 4 …辞書管理プログラム、3 5 …ユーザ入力情報管理プログラム、4 4 …コードデータ、4 5 …辞書データ

【書類名】 図面

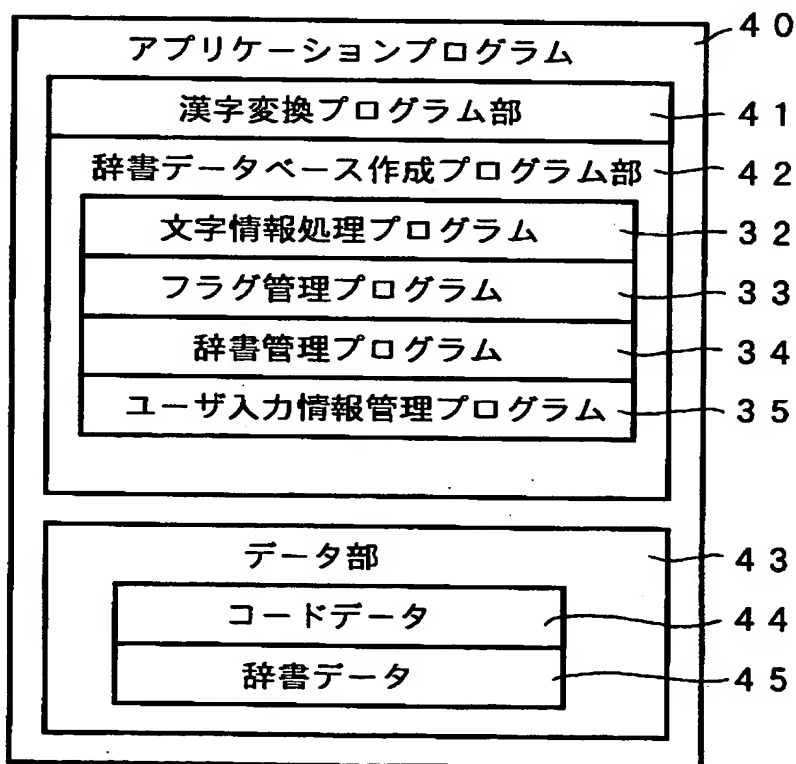
【図 1】



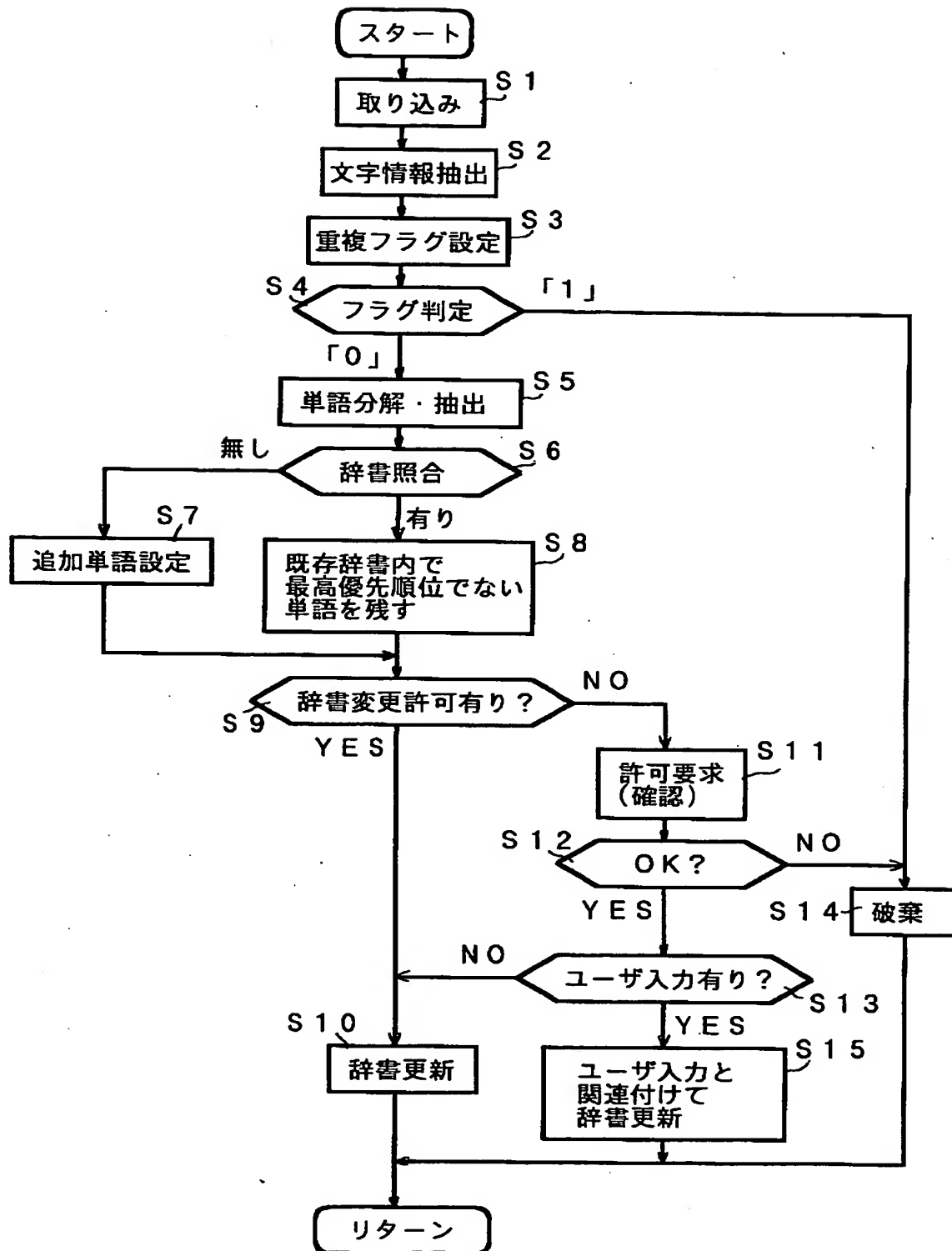
【図 2】



【図 3】



【図 4】



【書類名】 要約書

【要約】

【課題】 漢字変換システムの辞書データベースに対する新たな単語の追加及び更新や変換優先順位の設定を、ユーザによる複雑な操作を必要とせず、簡単且つ容易に実現可能とする。

【解決手段】 例えばユーザがインターネット上で検索、閲覧しているページやユーザがやりとりしている電子メール情報等をソース情報として取り込み（ステップ S 1）、そのソース情報から文字情報を抽出（ステップ S 2）し、その文字情報から単語情報を抽出（ステップ S 5）し、その単語情報を用いて辞書データベースを構築（ステップ S 6～ステップ S 10, S 15）する。

【選択図】 図 4

特 2000-366729

認定・付加情報

特許出願の番号	特願2000-366729
受付番号	50001551246
書類名	特許願
担当官	第七担当上席 0096
作成日	平成12年12月 4日

<認定情報・付加情報>

【提出日】	平成12年12月 1日
-------	-------------

次頁無

出 願 人 履 歴 情 報

識別番号 [395015319]

1. 変更年月日	1997年 3月31日
[変更理由]	住所変更
住 所	東京都港区赤坂7-1-1
氏 名	株式会社ソニー・コンピュータエンタテインメント